**Name:\_\_Sydney Wood\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**NOTE: Type your answers in the appropriate fields; please make answer fields larger as needed. Turn in a printed copy to your TAs during lab next Friday (October 19, 2018). Please note, assignments will lose 5% of the total possible points for each day they are late.**

*Conceptual Questions*

1. What are the formulas for standard deviation (i.e., standard versus alternative fomulae)? Describe in your own words what standard deviation describes in a distribution.

(Biased)σ or SD = Σ\* (x*i* – μ)2/N

(Unbiased) σ or SD= Σ\*(x*i* – μ)2 / (N-1) – this should be the default

Standard deviation is a variability statistic that indicates the average distance of each score from the mean. The standard deviation can be used to estimate the probability of which any specific score would occur.

\* this is the sum of all values in the sample but I couldn’t figure out how to type that notation: above ∑ = N, below ∑ = *i*=1

1. What expectations do we have given the central limit theorem (i.e., what are the claims of the central limit theorem)? In your own words explain what these expectations mean and why they make sense conceptually.

The central limit theorem states the occurrences of any thing be, it height, weight, etc. Tend to fluctuate, but naturally be pulled toward the mean. My favorite example of why this logically must be true is the concept of heritability of height. If the central limit theorem were false, we would live in a world of ity-bity ant people and giant people. Why? Because if height did not have a natural tendency to pull closer to the mean. Tall people would have even taller children and each generation after that would be taller and taller and taller. The same would happen for short people. However we do not live in a world of giants and ant people, instead children of tall people are more likely to be slightly shorter than their parents and short parents often have children who are taller than them. That same phenomenon is seen in most testable constructs.

As such this tendency to pull toward the mean create a pattern of scores that are naturally normally distributed. Normal distribution refers to the shape and statistical properties of the most common type of distribution. Normally Distributed constructs allow for robust probabilistic inferences. The probability inferences are the bases of statistics with regard to hypothesis testing.

*Computational Problems*

Use the same data set from last week, HW01Data.csv.

1. a. Create a histogram of weight. Create a unique figure title and label the x-axis.

b. Create a density plot of height. Create a unique figure title and label the x-axis. Add a solid black vertical line to the plot at the mean value of height. Add a dotted black vertical line to the plot at the median value of height. [Hint: look up the abline() function to learn about the argument for line types.]

Code/Syntax:

1. hist(hw02data$weight\_lbs, col = "deepskyblue", main = "Distribution of Participant Weights", xlab = "Weight in Pounds")
2. heightdensity <- density(hw02data$height\_in)

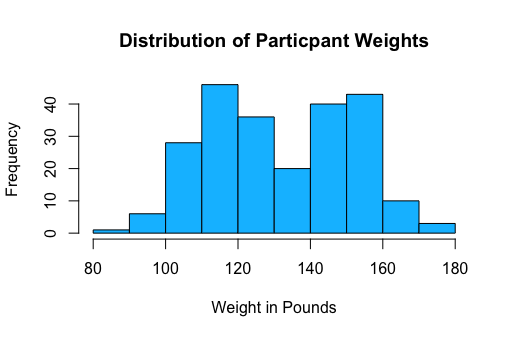
plot(heightdensity, main = "The Density of Participant Height", xlab = "Hieght in Inches")

abline(v = mean(hw02data$height\_in))

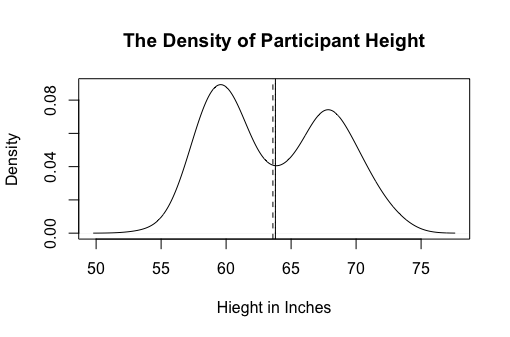
abline(v = median(hw02data$height\_in), lty = 2)

Answer:

a)



b)



1. Calculate the mean, median, variance, minimum, and maximum values of BMI for each unique condition of biological sex and education completed—you’ll need to calculate BMI again using the formula provided in HW01 (if you saved your syntax you can just copy and paste). Summarize these values in a table. Comment on how these groups compare on these statistics.

Code/Syntax: hw02data$bmi <- (hw02data$weight\_lbs/ (hw02data$height\_in^2))\* 703

descstats <- function(x) c(mean(x), median(x), var(x), min(x), max(x))

aggregate(hw02data$bmi~biosex\*ed\_cmplt, data = hw02data, descstats)

Answer:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1.  BMI Descriptive Statistics based on Gender and Level of Education Completed | | | | | | | |
| Gender | Level of Education Completed | BMI  Mean | BMI  Median | BMI  Variance | BMI  Min | BMI  Max | Comments |
| Female | High School | 22.73 | 22.73 | 8.64 | 18.15 | 26.78 | The means between groups are not drastically different. However, the spread between groups seems to vary depending on education achieved. In addition to the differences in spread seen within gender. There are also significant differences in the spread between genders with relation to education completed. If we had only looked at the means we would have assumed that there was little/no difference between groups, but looking at all statistics we can tell that the spread of the groups is very different. |
| 2 Year College | 23.01 | 22.37 | 3.91 | 20.28 | 25.97 |
| 4 Year College | 22.84 | 22.17 | 5.06 | 18.1 | 27.75 |
| Masters | 22.09 | 22.29 | 6.01 | 16.93 | 26.24 |
| PhD or Equivalent | 23.48 | 24.04 | 6.67 | 17.63 | 27.54 |
| Male | High School | 23.11 | 22.72 | 6.51 | 16.93 | 28.1 |
| 2 Year College | 22.42 | 22.17 | 5.76 | 18.43 | 30.23 |
| 4 Year College | 22.04 | 21.95 | 2.5 | 19.1 | 24.8 |
| Masters | 23.21 | 24.09 | 5.9 | 17.71 | 26.2 |
| PhD or Equivalent | 22.74 | 22.47 | 3.28 | 18.26 | 27.88 |

1. a. Identify the rows with the maximum and minimum BMI values. Report the row numbers and the maximum and minimum BMI values.

b. Standardize BMI values so that they have a mean of zero and standard deviation of one. Now identify the rows with the maximum and minimum standardized BMI values. What are the row numbers and what are the maximum and minimum standardized BMI values?

c. Standardize BMI values so that the mean is 100 and the standard deviation is 15. Identify the rows with the maximum and minimum standardized BMI values. What are the row numbers and what are the maximum and minimum standardized BMI values.

d. What conclusions can you draw about linear transformations of data based on your answers to the preceding questions (i.e., 5a-5c)?

Code/Syntax:

a) bmimax <- max(hw02data$bmi)

maxrow <- which(hw02data$bmi == bmimax)

bmimax

maxrow

bmimin <- min(hw02data$bmi)

minrow <- which(hw02data$bmi == bmimin)

bmimin

minrow

b) hw02data$zbmi <- scale(hw02data$bmi)

zbmimax <- max(hw02data$zbmi)

zmaxrow <- which(hw02data$zbmi == zbmimax)

zbmimax

zmaxrow

zbmimin <- min(hw02data$zbmi)

zminrow <- which(hw02data$zbmi == zbmimin)

zbmimin

zminrow

c) hw02data$newbmi <- (scale(hw02data$bmi)\*15 + 100)

newbmimax <- max(hw02data$newbmi)

newmaxrow <- which(hw02data$newbmi == newbmimax)

newbmimax

newmaxrow

newbmimin <- min(hw02data$newbmi)

newminrow <- which(hw02data$newbmi == newbmimin)

newbmimin

newminrow

d) no code

Answer:

a) maximum = 30.23 row = 63

Minimum = 16.84 row = 72

b) zmaximum = 3.23 row = 63

zminimum = -2.53 row = 72

c)newmaximum = 148.43 row = 63

new minimum = 61.99 row = 72

d) linear transformations change the actual values of each score, but it does not change the “relationship” of the scores compared to each other. If we were to graph all three bmi’s we would have the exact same shape just in different areas of the plot.